(19) RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE

PARIS

(11) N° de publication : **2 746 526**
(à n'utiliser que pour les commandes de reproduction)

(21) N° d'enregistrement national : **96 03690**

(51) Int Cl⁶ : G 06 F 12/16

(12) DEMANDE DE BREVET D'INVENTION A1

(22) **Date de dépôt** : 25.03.96.

(30) **Priorité** :

(43) **Date de la mise à disposition du public de la demande :** 26.09.97 Bulletin 97/39.

(56) **Liste des documents cités dans le rapport de recherche préliminaire :** *Se reporter à la fin du présent fascicule.*

(60) **Références à d'autres documents nationaux apparentés :**

(71) **Demandeur(s) :** *DIGITAL EQUIPMENT CORPORATION* — US.

(72) **Inventeur(s) :** GREEN RUSSELL J, DAVIES J CHRISTOPHER, PAXTON ALAN J et WHITAKER CHRISTOPHER.

(73) **Titulaire(s) :** .

(74) **Mandataire :** CABINET MALEMONT.

(54) PROCEDE POUR CONSERVER UNE BASE DE DONNEES A ORGANISATION TEMPORELLE ET SPATIALE.

(57) L'invention concerne un procédé pour conserver des informations dans une mémoire d'un système informatique, comprenant l'organisation des informations dans une première mémoire sous la forme d'une structure de données, la structure de données ayant un ordonnancement hiérarchique qui comprend un bas et un haut; et le stockage chronologique de la structure de données dans une seconde mémoire dans un ordre de bas en haut de l'ordonnancement hiérarchique afin de créer une base de données.

FR 2 746 526 - A1

2746526

Procédé pour conserver une base de données à organisation temporelle et spatiale

La présente invention concerne, d'une manière générale, des systèmes informatiques et, plus particulièrement, un procédé pour conserver en permanence une structure de données d'un système informatique sous la forme d'une base de données à organisation temporelle et spatiale.

Des structures de données sont utilisées pour organiser les informations traitées par les systèmes informatiques. D'une manière générale, les informations de la structure de données sont, pendant leur manipulation, stockées dans une mémoire à accès direct (RAM). Si la structure de données est trop importante pour être logée dans la mémoire RAM, elle peut également être organisée sous la forme d'une base de données et être stockée en permanence sur un dispositif de mémoire à disque à accès direct. A titre de sécurité, des copies de la base de données peuvent également être conservées sur un support de mémorisation à accès séquentiel bon marché, tel qu'une bande magnétique.

Les structures de données varient dans l'espace et dans le temps. Les structures de données traditionnelles organisent généralement la relation spatiale des informations pour gagner en efficacité de traitement. Dans la plupart des bases de données connues, un gain d'efficacité est réalisé grâce à un contrôle de la correspondance logique/physique des structures de données afin d'améliorer l'accès aux données. Par exemple, une base de données caractéristique peut contenir des enregistrements d'index et des enregistrements de données organisés hiérarchiquement sous la forme de noeuds d'un arbre binaire. Un emplacement connu stocke habituellement un pointeur indiquant un noeud de base de l'arbre. Le noeud de base contient des pointeurs indiquant d'autres noeuds d'index dans la hiérarchie. Les noeuds d'index de niveau inférieur comprennent des pointeurs indiquant les enregistrements de données. Les pointeurs ne sont en fait rien de plus que des adresses permettant au système de localiser physiquement les noeuds et les enregistrements de données.

Des codes sont associés aux enregistrements de données. Les codes identifient d'une manière unique des enregistrements individuels. Les codes, ou des séries de codes, peuvent également être associés aux enregistrements d'index afin de définir des chemins d'accès à des enregistrements de données particuliers. Ainsi, au cours du fonctionnement de la base de données, les codes peuvent servir à suivre sélectivement les pointeurs indiquant l'emplacement physique des enregistrements de données particuliers.

Bien que les structures de données à organisation spatiale puissent présenter des caractéristiques d'accès direct satisfaisantes tout en préservant un ordonnancement codé séquentiel, d'autres opérations risquent d'en pâtir. Dans la plupart des systèmes informatiques, une tâche fréquente et importante consiste à faire une copie de sauvegarde de la base de données sur un support amovible, par exemple. Pendant l'exécution de la copie de sauvegarde, la plupart des systèmes informatiques interdisent une modification de la base de données pour que la copie de sauvegarde représente une image conforme de la base de données à un moment particulier dans le temps. Pour les bases de données de très grande capacité, la période de temps pendant laquelle la base de données reste inaccessible pose un problème fonctionnel grave. Un autre problème rencontré dans les bases de données de l'art antérieur est qu'il est difficile de restaurer sélectivement des données à partir d'une copie de sauvegarde physique sans disposer d'un double support pour recevoir la totalité de la copie de sauvegarde pendant la restauration.

Pour limiter le temps d'indisponibilité de la base de données, certains systèmes informatiques exécutent la sauvegarde de manière incrémentielle. Dans une sauvegarde incrémentielle, seules les données modifiées depuis la dernière sauvegarde sont copiées sur le support de sauvegarde. Ainsi, une sauvegarde complète et toutes les sauvegardes incrémentielles suivantes peuvent servir à

restaurer la base de données après une défaillance du système.

Pour exécuter des sauvegardes incrémentielles, il faut connaître la relation temporelle des données. Cela signifie que la procédure de sauvegarde doit être capable d'identifier les données qui ont été modifiées depuis la dernière sauvegarde.

Toutefois, dans la plupart des bases de données à organisation spatiale traditionnelles, les données sont changées de place et l'évolution historique des données est perdue. Un journal conservé séparément comprenant des transactions de base de données horodatées peut être utilisé pour enregistrer les modifications de la base de données. Cependant, un journal augmente la complexité du système. De plus, une restauration sélective à partir d'un journal ne comportant qu'une organisation temporelle peut prendre beaucoup de temps car les relations spatiales requises des données récupérées doivent être reconstituées.

Un autre problème qui n'est habituellement pas résolu de manière efficace par les bases de données traditionnelles est de voir ou de sélectionner une partie de la base de données en fonction de contraintes temporelles. Une tâche caractéristique d'une base de données peut consister à identifier des données qui satisfassent non seulement des critères de sélection logique prédéterminés, mais également des critères de sélection temporelle comme, par exemple, la localisation de tous les enregistrements qui ont été modifiés au cours des dernières vingt-quatre heures. Dans la plupart des systèmes dans lesquels des horodateurs sont conservés avec les enregistrements de données, il est habituellement nécessaire, premièrement, de lire tous les enregistrements qui qualifient logiquement les données et, deuxièmement, de qualifier à nouveau les données en fonction des contraintes temporelles. A titre de variante, si un journal est utilisé, les deux étapes de l'opération sont habituellement exécutées dans un ordre inverse. Dans la plupart des bases de données, il est impossible de sélectionner des données en utilisant

simultanément des contraintes à la fois spatiales et temporelles.

D'une manière générale, les bases de données à organisation spatiale traditionnelles se prêtent mal à des opérations dépendantes du temps, et des opérations spatiales sur des structures de données à organisation temporelle sont également difficiles.

Compte tenu des inconvénients des bases de données de l'art antérieur, la présente invention a pour but de proposer une organisation de la structure de données d'une base de données, qui permette de réaliser d'une manière efficace des opérations fonction aussi bien du temps que de l'espace.

Pour atteindre ce but, la présente invention propose un procédé pour conserver des informations dans une mémoire d'un système informatique, comprenant l'organisation des informations dans une première mémoire sous la forme d'une structure de données, la structure de données ayant un ordonnancement hiérarchique qui comprend un bas et un haut ; et le stockage chronologique de la structure de données dans une seconde mémoire dans un ordre de bas en haut de l'ordonnancement hiérarchique afin de créer une base de données.

Au cours du fonctionnement d'un système informatique, des enregistrements d'index et des enregistrements de données d'une structure de données sont organisés d'une manière hiérarchique, les enregistrements d'index se situant à un niveau hiérarchique plus élevé que les enregistrements de données. Pendant que la structure de données est manipulée, les enregistrements de données concernés sont stockés dans une mémoire à accès direct. Les modifications apportées à la structure de données sont inscrites chronologiquement dans une mémoire à disque dans un ordre hiérarchique de bas en haut en vue d'un stockage permanent sous la forme d'une base de données. Une copie de sauvegarde de la base de données est exécutée par stockage des enregistrements de données et d'index de la base de données dans un ordre de haut en bas sur un support de sauvegarde à lecture séquentielle. Une vue

temporelle et spatiale d'une partie de la base de données peut être obtenue par un accès aux enregistrements d'index et aux enregistrements de données de la base de données ou au support de sauvegarde suivant l'ordre de haut en bas.

Les enregistrements de données et les enregistrements d'index sont transférés entre la mémoire à accès direct et la mémoire à disque sous la forme de segments. Un numéro de segment logique unique est associé à chaque segment. Des numéros de segments logiques croissant d'une manière monotone sont attribués aux segments lorsque ceux-ci sont inscrits dans la mémoire à disque. Un emplacement initial stocke le numéro de segment logique du segment ayant le numéro le plus élevé, inscrit dans la mémoire à disque et l'heure de l'inscription du segment ayant le numéro de segment logique le plus élevé.

Au cours d'une opération de sauvegarde, les segments sont lus dans un ordre chronologique inverse à partir de la mémoire à disque, et sont inscrits dans l'ordre chronologique inverse sur le support de sauvegarde. Le support de sauvegarde offre ainsi une disposition de haut en bas des enregistrements d'index et des enregistrements de données pour permettre d'effectuer une restauration temporelle d'enregistrements d'index et de données sélectionnés, en lisant le support de sauvegarde dans un sens direct unique.

Ce qui précède ressortira plus clairement de la description détaillée suivante d'un mode de réalisation préféré de la présente invention, donnée à titre d'exemple nullement limitatif en référence aux dessins annexés dans lesquels :

La figure 1 est un schéma fonctionnel d'un système informatique comportant une structure de données organisée conformément aux principes de l'invention ;

La figure 2 est un schéma fonctionnel d'une structure de données hiérarchique à un premier moment dans le temps ;

La figure 3 est un schéma fonctionnel de la structure de données de la figure 2 à un moment ultérieur dans le temps ;

La figure 4 représente une mise en correspondance de la

structure de données de la figure 2 avec un support de mémorisation visible sur la figure 1, selon l'invention ;

La figure 5 est un schéma fonctionnel d'un segment du support ;

La figure 6 représente une mise en correspondance de la structure de données de la figure 3 avec le support de mémorisation ;

La figure 7 est un organigramme d'une procédure de sauvegarde ;

La figure 8 représente un ensemble de sauvegarde créé par la procédure de la figure 7 ; et

La figure 9 est un organigramme d'un traitement de restauration d'un ensemble de sauvegarde.

En référence aux dessins et en particulier à la figure 1, celle-ci représente un système informatique 100 comprenant une unité centrale de traitement (CPU) 110, une mémoire 120, et un sous-système d'entrée/sortie (I/O) 130 reliés les uns aux autres par un bus de communication 140. Un support de mémorisation de grande capacité à accès direct 150 comprenant, par exemple, une ou plusieurs unités de disques est relié au sous-système I/O 130 par un bus d'entrée/sortie 145. Une partie de la mémoire 120 peut être attribuée à une antémémoire sur disque 121 pour tirer profit de la proximité des informations traitées par le système 100. Le système 100 peut également comprendre un système d'entraînement de bande 155 pour stocker une copie des informations traitées sur un support amovible et à accès séquentiel comme, par exemple, une bande magnétique 156. Les informations stockées sur le support du dispositif de mémorisation 150 peuvent être organisées sous la forme d'une base de données 160. La base de données 160 organise les informations ou les "données" pour faciliter un accès et offrir une plus grande fiabilité.

Au cours du fonctionnement du système informatique 100, des parties de la base de données 160 sont mises en mémoire dans l'antémémoire sur disque 121 pour être traitées par la CPU 110. Une fois que les données ont été traitées, les données modifiées sont réécrites sur le disque 150 en vue

d'un stockage permanent.

La figure 2 représente, sous la forme d'un graphique directionnel, la structure logique de la base de données 160 à un moment T1. Dans le mode de réalisation pris à titre d'exemple, la base de données 160 est organisée logiquement d'une manière hiérarchique sous la forme d'un "arbre" inversé 200. L'arbre 200 comprend un noeud d'index de niveau haut ou un noeud "de base" 210, des noeuds d'index de niveau intermédiaire 220, et des noeuds de données de niveau bas 230, parfois appelés "noeuds feuilles". Pour faciliter la description du mode de réalisation de l'invention pris à titre d'exemple, les noeuds portent des références A à G. Les noeuds d'index (A, B et C) contiennent des enregistrements d'index, et les noeuds de données (D, E, F et G) contiennent des enregistrements de données. On comprendra que la hiérarchie peut comprendre de multiples niveaux de noeuds d'index et que chaque noeud d'index peut contenir plusieurs enregistrements d'index.

Le réseau qui maintient la cohérence de la structure 160 est formé par les pointeurs 201 à 207. Par exemple, le pointeur 201 qui localise la structure 160 est habituellement stocké au niveau d'un emplacement "initial" (H) connu 200. Au cours du fonctionnement de la base de données, un accès physique aux enregistrements de données est possible en suivant les pointeurs 201 à 207, par exemple, les enregistrements d'index, en fonction d'un critère de sélection prédéterminé. Par conséquent, une adresse 211 et un code 212 sensible à un critère sont associés à chaque pointeur ou à chaque enregistrement d'index de l'arbre. D'une manière caractéristique, le code 212 peut servir à identifier d'une manière unique un enregistrement de données particulier. Ce type d'organisation de base de données peut être utilisé pour des systèmes de fichiers complexes et des bases de données relationnelles.

La figure 3 représente la base de données 160 de la figure 2, selon la présente invention, à un moment ultérieur T2. Des modifications apportées aux noeuds de données peuvent

avoir nécessité des changements correspondants dans les noeuds d'index. D'une manière générale, les modifications sont indiquées par le symbole prime ('). Par exemple, les modifications peuvent avoir changé le contenu ou la taille d'un enregistrement de données ou d'un enregistrement d'index en nécessitant éventuellement un repositionnement physique de celui-ci. Dans la plupart des bases de données traditionnelles, l'évolution temporelle de la base de données est perdue étant donné que les enregistrements d'index et de données sont en général changés de place. Cela signifie que les informations de données et d'index périmées de la base de données sont en général écrasées.

En revanche, dans la base de données 160 de l'invention, visible sur la figure 3, l'évolution historique de la base de données est, d'une manière générale, conservée. En d'autres termes, il est possible conformément à l'invention de récupérer, comme cela sera décrit d'une manière plus détaillée ci-après, les états de la base de données dans le temps. La conservation de l'évolution de la base de données 160 signifie que le support physique qui stocke les enregistrements d'index et les enregistrements de données ne doit pas être écrasé, sauf dans le cas particulier où les données sont trop anciennes ou trop "vieilles" pour être d'une quelconque utilité immédiate. Dans ce cas, les vieilles données peuvent être archivées pour permettre de récupérer de la place sur le support physique.

Dans l'exemple représenté sur la figure 3, la modification de la base de données 160 a nécessité d'apporter des changements aux noeuds de données et d'index. Les noeuds modifiés sont, d'une manière générale, indiqués par les références A', B' et E'. On notera que la structure de données initiale des noeuds A à G reste préservée. Le noeud modifié A' possède des pointeurs 202' et 203' qui adressent respectivement les noeuds B' et C, tandis que le noeud d'index modifié B' possède un pointeur 205' destiné à localiser le noeud de données modifié E'.

Pour refléter le fait que la base de données 160 possède

maintenant en fait deux noeuds de base, tels que les noeuds A et A', par exemple, l'emplacement initial H 200 a été modifié pour comprendre deux pointeurs 201 et 201' respectivement associés aux temps T1 et T2. Ainsi, à n'importe quel moment ultérieur, il est possible d'examiner l'état de la base de données 160 au temps T1 en suivant le pointeur 201 à partir de l'emplacement initial 200, et d'obtenir l'état de la base de données 160 au temps T2 en suivant le pointeur 201'. On notera qu'une vue de la base de données au temps T2 en suivant, par exemple, le pointeur 201' fait apparaître les noeuds de données D, E', F, et G, mais pas le noeud E. Il est clair maintenant que la base de données 160 peut continuer à évoluer au fur et à mesure que des enregistrements de données et d'index sont modifiés par la création de pointeurs supplémentaires dans l'emplacement initial 200 et que, surtout, l'évolution chronologique de la base de données 160 dans le temps est conservée. Précisément, comme on peut le voir, l'évolution de la base de données progresse du bas de la figure vers le haut.

La figure 4 représente un mode de réalisation proposé à titre d'exemple d'une mise en correspondance de la structure de données 160 avec un support physique à l'aide d'un réseau à la fois temporel et spatial. Sur la figure 4, le numéro de référence 400 désigne globalement la partie du support physique du disque 150 de la figure 1, qui a été attribuée au stockage de la base de données 160. L'emplacement initial 200 qui stocke l'adresse du noeud de base A 210 est connu. D'une manière générale, cela signifie que les seules informations qui sont stockées à un emplacement physique fixe de la base de données 160 sont les informations stockées au niveau de l'emplacement initial 200.

Dans la base de données 160, les noeuds, tels que les enregistrements de données et d'index, par exemple, de l'arbre 200 sont inscrits chronologiquement sur le support 400 dans un ordre de bas en haut. Les noeuds de données de niveau bas 230 sont inscrits, en termes de temps, avant les noeuds d'index de niveau intermédiaire 220 dont les pointeurs

sont dirigés sur eux. Le noeud de base de niveau haut 210 est inscrit en dernier. Ce classement temporel des données inscrites se reflète dans l'ordonnancement de gauche à droite des noeuds de la figure 4 ou dans la disposition de bas en haut de la figure 3.

Dans un mode de réalisation préféré, le classement chronologique est conservé grâce au fait que les données sont stockées à des adresses "logiques" progressivement plus élevées du support. Ainsi, on sait que des données inscrites à des adresses logiques inférieures sont plus anciennes que des données inscrites à des adresses logiques supérieures. De même, des données situées à un niveau supérieur de la relation hiérarchique auront également des adresses logiques plus élevées que des données situées à un niveau inférieur.

L'ordonnancement des adresses logiques s'obtient, en partie, en inscrivant chronologiquement les données sur le support 400 sous la forme de segments comme, par exemple, le premier segment 1 et le dernier segment 500. Un segment constitue l'unité de transfert de données entre l'antémémoire sur disque 121 et le disque 150 de la figure 1. Les segments peuvent occuper un ou plusieurs blocs de support discrets du disque 150. Si les blocs de support occupés par un segment particulier sont disposés physiquement les uns à côté des autres, les performances de lecture et d'écriture peuvent être améliorées. La taille de chaque segment se situe d'une manière caractéristique dans la plage de 100 kilo-octets à 10 mégaoctets.

La figure 5 représente en détail l'un de ces segments, par exemple, le segment 500. Un numéro de segment logique unique (LSN) 510 est associé à chaque segment 500. Les numéros de segments logiques 510 peuvent être attribués aux segments dans un ordre croissant d'une manière monotone au fur et à mesure que les segments sont inscrits sur le support 400.

Par conséquent, en plus d'identifier d'une manière unique les segments, les numéros de segments logiques LSN indiquent également l'âge relatif des segments ou l'ordre

chronologique dans lequel les segments ont été inscrits sur le support 400. La valeur du LSN d'un segment augmente à chaque fois qu'un nouveau segment est enregistré. Un segment ayant un LSN supérieur est toujours inscrit après, en termes de temps, un segment ayant un LSN inférieur. Une fois qu'un LSN est attribué à un segment, il ne change jamais, même si l'emplacement physique du segment varie. En fait, les numéros de segments logiques peuvent servir de pointeurs dans l'espace et dans le temps. Les numéros de segments logiques indiquent non seulement l'emplacement des données mais également leur âge relatif.

La mise en correspondance logique/physique entre les numéros de segments logiques et le support physique peut être réalisée au moyen d'un index de segments (SI) 550. L'index de segments 550 peut être stocké séparément sur le support 400 sous la forme d'une liste, par exemple. La tête de la liste peut correspondre au dernier segment inscrit, l'entrée suivante dans la liste pouvant représenter le segment qui le précède immédiatement, et ainsi de suite. L'adresse de l'index de segments 550 peut être stockée au niveau de l'emplacement initial 200. Le SI 550 établit une correspondance de un pour un entre le numéro de segment logique 510 et une adresse de segment physique (PSA) 555 du segment. L'index de segments 550 permet un agencement optimal des segments sur le support ainsi qu'un repositionnement des segments sans perturber les relations logiques, spatiale et temporelle, par exemple, entre les segments.

Lorsque toutes les données d'un segment particulier ont été modifiées, le segment est considéré comme "vieux" et n'est plus accessible à partir du noeud de base de la base de données. Le support physique occupé par de vieux segments peut être recyclé pour accueillir de nouvelles données. Lorsqu'un segment est recyclé, l'index de segments 550 est marqué pour indiquer que le numéro de segment logique 510 associé au segment devenu vieux n'est plus en service. Le numéro de segment logique 510 n'est bien entendu jamais recyclé pour conserver son caractère unique, seul le support

étant recyclé.

Bien que l'unité d'accès physique fondamentale de la base de données 160 soit le segment, un accès logique aux données est possible au niveau des enregistrements, par exemple, au niveau des enregistrements de données 520 et des enregistrements d'index 530. Par conséquent, une adresse de décalage d'enregistrement (ROA) 540 est associée à chaque enregistrement pour indiquer l'emplacement relatif de l'enregistrement à l'intérieur du segment et est exprimée, par exemple, sous la forme d'une adresse d'un octet. Ainsi, si un pointeur de "noeud" est exprimé sous la forme d'une concaténation du LSN 510 et de la ROA 540, telle que "ROA_LSN", par exemple, l'emplacement sur le support 400, où les données concernées sont stockées peut être déterminé facilement. On notera que les pointeurs décrits ici possèdent, contrairement aux pointeurs de l'art antérieur, des attributs spatial, temporel et physique. Les pointeurs indiquent l'endroit où les informations sont situées sur le support, quand les données ont été créées et comment y accéder.

La figure 6 représente l'état de la base de données 160 au temps T2 correspondant à la figure 3. Comme cela a été décrit précédemment, au cours de l'évolution de la base de données 160 entre le temps T1 et le temps T2, les noeuds A, B et E, c'est-à-dire leurs enregistrements associés, ont été modifiés. Les noeuds modifiés A', B' et E' sont inscrits sur le support dans un ordre de bas en haut. Par conséquent, pour voir l'état de la base de données à n'importe quel moment dans le temps, il suffit de déterminer le LSN du noeud de base à ce moment donné et de suivre les pointeurs en descendant le long de l'arbre. Cette vue est appelée un "instantané" de la base de données.

De temps en temps, il est nécessaire de faire une copie des informations stockées dans la base de données à des fins de sauvegarde, afin de faire en sorte que les informations ne soient pas perdues à cause d'une défaillance. Une défaillance peut être due aux conditions environnantes, à une panne de

matériel ou de logiciel, ou à des défauts d'origine fonctionnelle.

Conformément aux principes de l'invention, la sauvegarde, quelle soit totale ou incrémentielle, peut être exécutée sensiblement en direct sans interruption majeure du fonctionnement normal. L'invention permet également une récupération au moment opportun des informations perdues, de sorte qu'un fonctionnement normal peut reprendre rapidement après une défaillance. En outre, les informations peuvent être récupérées sélectivement sans qu'il soit nécessaire de restaurer préalablement la totalité de la copie de sauvegarde.

Dans le mode de réalisation préféré de la présente invention, la sauvegarde est réalisée, en partie, par la saisie d'un état cohérent de lecture seule de la base de données à l'aide d'un instantané. Une fois que l'instantané a été réalisé, un accès et une modification de la base de données peuvent se poursuivre pendant que la copie de sauvegarde est exécutée. Au cours de l'exécution de la copie de sauvegarde, le support qui stocke les données de l'instantané ne peut pas être récupéré puisque la sauvegarde conserve un état exact du support. Lorsque les segments sont copiés sur le support de sauvegarde, l'espace peut être récupéré.

En termes plus simples, l'exécution d'un instantané inscrit toutes les données de la base de données 160 sur le support 400. D'une manière caractéristique, au cours du fonctionnement normal de la base de données 160, les données qui font l'objet d'accès fréquents, telles que les enregistrements d'index, sont habituellement conservées dans l'antémémoire sur disque 121 de la mémoire 120 de la figure 1. Par conséquent, l'exécution d'un instantané signifie le transfert des données modifiées ou "altérées" de l'antémémoire sur disque 121 sur le disque 150, en segments ou en unités de segments partiels, comme cela a été décrit précédemment.

Lorsque les données sont inscrites, les pointeurs sont

mis à jour, et les enregistrements sont inscrits dans un ordre temporel et spatial de bas en haut. Etant donné que l'antémémoire sur disque 121 ne stocke d'une manière caractéristique qu'une fraction de la base de données totale, et que toutes les données conservées dans l'antémémoire 121 ne sont pas modifiées ou altérées, l'instantané peut être fait très rapidement. Une fois que la totalité de la base de données réside sur le support, le numéro de segment logique 510 du dernier segment inscrit est stocké au niveau de l'emplacement initial 200. Ce LSN 510 définit le point de départ pour fournir une vue cohérente de la base de données à ce moment dans le temps, approprié pour une sauvegarde. Dès que l'emplacement initial 200 a été mis à jour à l'aide d'un nouvel horodateur et d'un nouveau pointeur de niveau haut, des modifications en temps réel de la base de données 160, c'est-à-dire des opérations normales, peuvent reprendre. L'invention garantit que les modifications apportées ultérieurement à la base de données 160 ne perturberont pas les états antérieurs de la base de données.

La figure 7 montre une vue d'ensemble d'un traitement 700 qui peut être utilisé pour sauvegarder et restaurer la base de données 160 après une défaillance. Au cours de l'étape 710, un "instantané" de lecture seule 180 de la base de données 160 est exécuté. Par essence, l'instantané contraint la totalité de la base de données à passer de l'antémémoire 121 sur le support de mémorisation, et détermine le LSN le plus élevé au moment de l'instantané. Au cours de l'étape 720, l'instantané est copié sur le support de sauvegarde, c'est-à-dire sur la bande magnétique 156, par exemple, sous la forme d'un "ensemble de sauvegarde" 800. La copie de l'instantané signifie la copie de tous les segments qui peuvent être localisés en suivant les pointeurs à partir du noeud de base au moment de l'instantané. Au cours de la sauvegarde, les "vieux" segments ne sont de préférence pas transférés sur le support de sauvegarde. Au lieu de parcourir logiquement l'arbre pour localiser les données qui doivent être copiées, une copie physique peut être exécutée au cours

de la sauvegarde à l'aide des informations stockées dans l'index de segments 550.

Après une défaillance, les informations de la base de données peuvent être récupérées dans une base de données "restaurée" 161 à partir de l'ensemble de sauvegarde 800 au cours de l'étape de restauration 730. Etant donné que la base de données 160 est structurée chronologiquement, toutes les informations qui, en terme de temps, précèdent l'instantané font partie de l'ensemble de sauvegarde 800, tandis que les informations inscrites sur le support après l'instantané n'en font pas partie. Ainsi, la base de données 160 peut rester accessible pour un accès en direct pendant la création de l'ensemble de sauvegarde 800 au cours de l'étape 720.

Conformément à l'invention, l'ensemble de sauvegarde 800 peut être soit un ensemble de sauvegarde total soit un ensemble de sauvegarde incrémentiel. L'ensemble de sauvegarde total comprend tous les segments de la base de données 160 jusqu'au moment de l'instantané. L'ensemble de sauvegarde incrémentiel ne comprend que les segments qui ont été inscrits depuis la dernière sauvegarde. En référence aux figures 3 et 6 par exemple, une sauvegarde incrémentielle au temps T2 ne copierait que les segments qui stockent des données associées aux noeuds modifiés A', B' et E'.

La figure 8 représente la structure de l'ensemble de sauvegarde 800 selon la présente invention. La structure définit la manière dont les informations de l'instantané 180 sont stockées sur le support de sauvegarde cible, c'est-à-dire sur la bande magnétique 156, par exemple. L'ensemble de sauvegarde 800 comprend des métadonnées 810 et les segments, par exemple les segments 500 à 1. Les métadonnées 810 identifient simplement l'ensemble de sauvegarde 800 et décrivent les caractéristiques de la base de données 160 dont elles constituent une sauvegarde. Les métadonnées 810 comprennent la quantité de segments logiques la plus importante au moment de l'instantané, l'heure à laquelle l'instantané a été créé et toutes les autres informations utiles pour identifier la sauvegarde. Les caractéristiques de

la base de données peuvent comprendre des caractéristiques telles que la taille des segments et des facteurs de blocage de support. Les métadonnées 810 sont utilisées au cours de l'étape de restauration 730 de la figure 7 pour configurer la base de données restaurée 161. Les métadonnées 810 comprennent également des informations mettant en correspondance les LSN avec leur emplacement physique sur le support à bande. Ces informations sont semblables à celles qui sont conservées dans l'index de segments 550 du support à disque. Ces informations de mise en correspondance peuvent servir, pendant une opération de récupération, à lire des segments spécifiques en sautant des segments intermédiaires.

Conformément aux principes de la présente invention, les segments de l'ensemble de sauvegarde 800 sont inscrits sur le support d'ensemble de sauvegarde dans l'ordre de haut en bas. En d'autres termes, l'ordonnancement des segments dans l'ensemble de sauvegarde est inversé en termes d'espace et de temps par rapport à l'ordonnancement des segments dans la base de données. Dans l'ensemble de sauvegarde 800, les segments ayant un numéro de segment logique supérieur précèdent les segments ayant un numéro de segment logique inférieur.

Cela signifie que la structure d'arbre hiérarchique de la base de données 160 est inscrite sur le support de sauvegarde par une lecture des segments de la base de données 160 dans un ordre chronologique inverse, c'est-à-dire dans un ordre décroissant des LSN. Ainsi, le noeud de base le plus récent est inscrit dans un segment de l'ensemble de sauvegarde 800 précédant le segment contenant les noeuds d'index, et les segments des noeuds d'index sont inscrits avant les segments de noeuds de données. Si le support de sauvegarde est une bande magnétique, la lecture et l'écriture des segments peuvent être imbriquées pour permettre un "déroulement continu" de la bande à une vitesse maximale.

Le stockage des segments dans l'ordre de haut en bas permet de récupérer complètement et sélectivement les informations en un seul passage de lecture du support de

sauvegarde, en parcourant la structure hiérarchique de haut en bas. Ceci constitue un avantage considérable si le support est une bande magnétique, car le repositionnement de la bande est notoirement inefficace si celle-ci est soumise à des inversions fréquentes de sens.

Exception faite de leur ordonnancement logique inversé, les segments de l'ensemble de sauvegarde sont une image exacte des segments du volume. Par exemple, au cours de l'étape 720 d'exécution de l'ensemble de sauvegarde, de la figure 7, les segments sont traités d'une manière opaque, sans déchiffrage ni traitement de leur contenu.

La figure 9 représente les étapes qui peuvent être exécutées pour restaurer la base de données. Au cours du traitement de restauration, le support 161 est invisible pour le reste du système 100. Au cours de l'étape 910, les métadonnées 810 sont lues à partir de l'ensemble de sauvegarde 800. Au cours de l'étape 920, le support destiné à recevoir la base de données restaurée est attribué et initialisé en fonction des métadonnées 810. La lecture des métadonnées 810 permet de récupérer également l'index des segments.

Au cours de l'étape 930, les segments sont lus à partir de l'ensemble de sauvegarde 800 dans l'ordre, par exemple, l'ordre chronologique inverse, de haut en bas, de celui dans lequel ils ont été stockés. Pour une récupération totale, les segments peuvent être copiés d'une manière transparente sans aucun autre traitement. Les numéros de segments logiques ne sont pas modifiés bien que leurs emplacements physiques puissent avoir changé par rapport à ceux où ils se trouvaient sur le support initial. Cela signifie qu'il peut être nécessaire de recréer l'index de segments 550 au cours de la restauration en fonction de la mise en correspondance logique/physique utilisée pendant la restauration. Si la lecture est imbriquée avec l'écriture, la bande peut se dérouler en continu à une vitesse maximale. Finalement, au cours de l'étape 940, le traitement de sauvegarde peut signaler au système que la base de données restaurée 161 est

prête à être utilisée.

Bien que le traitement de restauration 900 ait été optimisé pour réaliser une sauvegarde et une restauration totales, une récupération partielle des enregistrements de données est également facilitée. De plus, grâce à l'ordonnancement de haut en bas des informations dans l'ensemble de sauvegarde 800, il est possible de procéder à une récupération sélective en effectuant un seul passage à travers le support de sauvegarde. Il suffit simplement que des informations de restauration sélective 935 concernant les données qui doivent être récupérées soient fournies à l'étape 930. Au fur et à mesure que l'ensemble de sauvegarde 800 est parcouru, les enregistrements sont récupérés d'une manière optimale par la seule visite des segments qui stockent des informations importantes, les autres segments étant sautés.

L'ordonnancement unique de la base de données 160 selon l'invention offre des avantages pratiques. Un "instantané" de lecture seule de la base de données peut être créé très facilement et très rapidement à un moment donné dans le temps. L'instantané peut être utilisé pour réaliser une sauvegarde en direct sans gêner les opérations en cours de la base de données. En outre, la sauvegarde peut être réalisée sans autre interprétation des informations copiées pour accélérer le processus. En d'autres termes, la sauvegarde peut être une copie physique très rapide du support en ligne sur le support de sauvegarde.

D'autre part, l'organisation proposée par la présente invention permet également une véritable sauvegarde incrémentielle. Dans une sauvegarde incrémentielle, seules les parties de la base de données qui ont été modifiées depuis la dernière sauvegarde sont copiées sur le support de sauvegarde. Dans les bases de données traditionnelles, les sauvegardes incrémentielles peuvent consommer des ressources importantes du système, étant donné qu'il n'est pas toujours facile de déterminer les données qui ont été modifiées depuis la dernière sauvegarde. En outre, l'organisation de base de données de la présente invention permet une restauration

logique de parties de la base de données après une défaillance. Au cours de la restauration logique, seules des informations spécifiées sont restaurées de la copie de sauvegarde sur le support en ligne.

Un autre mode de réalisation de la présente invention permet de voir uniquement les enregistrements de données de la base de données qui ont été modifiés au cours d'un intervalle de temps relatif spécifique. Il est également possible de ne voir que les enregistrements qui sont qualifiés par des critères de sélection logique prédéterminés, comme par exemple une combinaison booléenne d'un ou de plusieurs codes.

Dans une base de données traditionnelle, des opérations temporelles ne peuvent être exécutées que si un journal séparé est conservé ou si les enregistrements contiennent des horodateurs de transactions. Dans les deux cas, un balayage très long de la totalité de la base de données ou du journal est nécessaire pour localiser uniquement les enregistrements qualifiés par les deux restrictions de "code" et de "temps".

En référence à nouveau aux figures 3 et 6, les instantanés aux temps T1 et T2 sont pris, par exemple, à vingt-quatre heures d'intervalle. Il est souhaitable de ne voir que les enregistrements de données sélectionnés qui ont été modifiés au cours de l'intervalle des dernières vingt-quatre heures. Par conséquent, une vue de la base de données 160 commence au niveau du noeud de base actuel A' 210. Selon le critère de sélection logique, par exemple un code, la structure de l'index est parcourue et seuls les enregistrements de données qualifiés, stockés dans les segments inscrits au cours de l'intervalle de T1 à T2 sont extraits.

Grâce à la structure de l'invention, représentée sur la figure 6, il est possible de terminer le parcours dès que le premier segment inscrit après le temps T1 a été lu. Il est inutile de continuer à parcourir la base de données. Cela signifie qu'il n'est pas nécessaire de suivre les pointeurs qui coupent la limite de temps T1, comme par exemple les

pointeurs 203' et 204', puisque l'on sait que les données concernées ont été modifiées préalablement à l'intervalle de temps considéré. Ceci constitue une amélioration significative des traitements utilisés dans les bases de données traditionnelles.

Bien que la description précédente ait porté sur un mode de réalisation préféré de la présente invention, celle-ci n'est bien entendu pas limitée à l'exemple particulier décrit et illustré ici, et l'homme de l'art comprendra aisément qu'il est possible d'y apporter de nombreuses variantes et modifications sans pour autant sortir du cadre de l'invention.

REVENDICATIONS

1. Procédé pour conserver des informations dans une mémoire d'un système informatique (100), caractérisé en ce qu'il comprend l'organisation des informations dans une première mémoire sous la forme d'une structure de données, la structure de données ayant un ordonnancement hiérarchique qui comprend un bas et un haut ; et le stockage chronologique de la structure de données dans une seconde mémoire (150) dans un ordre de bas en haut de l'ordonnancement hiérarchique afin de créer une base de données (160).

2. Procédé selon la revendication 1, caractérisé en ce qu'il comprend également l'introduction d'une partie de la base de données dans la première mémoire ; la modification de la partie de la base de données ; et le stockage chronologique de la partie de la base de données modifiée dans la seconde mémoire (150) dans l'ordre de bas en haut afin de mettre à jour la base de données (160).

3. Procédé selon la revendication 2, caractérisé en ce qu'il comprend également la lecture de la base de données (160) à partir de la seconde mémoire (150) dans un ordre chronologique inverse et de haut en bas ; et l'inscription de la base de données dans une troisième mémoire (155) dans l'ordre chronologique inverse et de haut en bas afin de créer une sauvegarde de la base de données.

4. Procédé selon la revendication 3, caractérisé en ce qu'il comprend également la détection d'une défaillance dans la seconde mémoire ; la lecture de la structure de données à partir de la troisième mémoire dans l'ordre chronologique inverse et de haut en bas ; et l'inscription de la structure de données dans la seconde mémoire dans l'ordre chronologique et de bas en haut afin de restaurer la base de données.

5. Procédé selon la revendication 1, caractérisé en ce que la structure de données hiérarchique contient des enregistrements d'index et des enregistrements de données, et en ce qu'il comprend également l'inscription des enregistrements d'index et des enregistrements de données sous la forme de plusieurs segments, les segments ayant des

numéros de segments croissants d'une manière monotone selon l'ordonnancement chronologique et hiérarchique de la base de données (160).

6. Procédé selon la revendication 3, caractérisé en ce qu'il comprend également le stockage d'un premier temps (T1) associé à la création de la sauvegarde de la base de données (160) ; la modification de la base de données jusqu'à un second temps (T2), second temps qui est relativement postérieur au premier temps ; la lecture d'une partie incrémentielle de la base de données à partir de la seconde mémoire dans l'ordre chronologique inverse et de haut en bas en commençant au second temps pour terminer au premier temps ; et l'inscription de la partie incrémentielle dans une quatrième mémoire dans l'ordre chronologique inverse et de haut en bas afin de créer une sauvegarde incrémentielle.

7. Procédé selon la revendication 6, caractérisé en ce qu'il comprend également la détection d'une défaillance dans la seconde mémoire ; la lecture, dans l'ordre chronologique inverse et de haut en bas, de la structure de données à partir de la troisième mémoire et de la partie incrémentielle à partir de la quatrième mémoire ; et l'inscription de la structure de données et de la partie incrémentielle dans la seconde mémoire dans l'ordre chronologique et de bas en haut afin de restaurer la base de données.

8. Procédé selon la revendication 1, caractérisé en ce qu'il comprend également la fourniture d'un critère de sélection logique et d'un critère de sélection temporelle, le critère de sélection temporelle ayant un premier temps et un second temps chronologiquement postérieur ; la lecture de la structure de données dans un ordre chronologique inverse en commençant au second temps pour terminer au premier temps ; et, pendant la lecture de la structure de données dans l'ordre chronologique inverse, la sélection simultanée de données en fonction du critère de sélection logique dans un ordre de haut en bas.

2746526

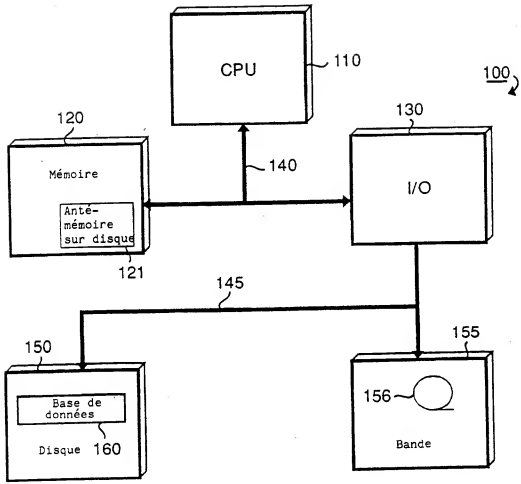

Fig. 1

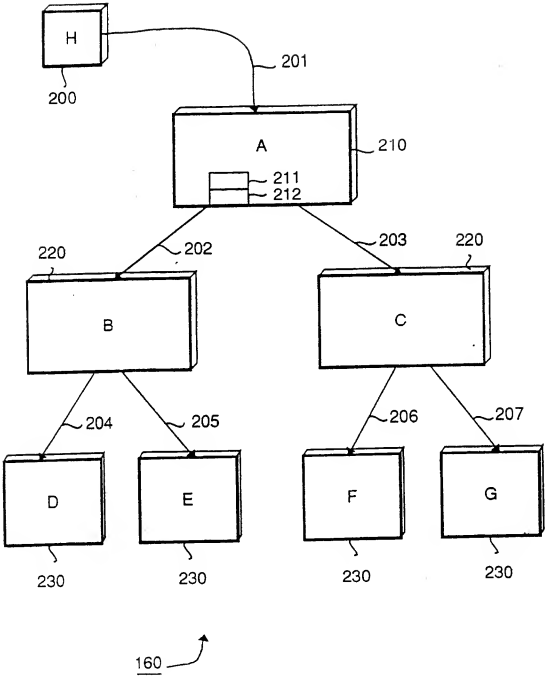
H
200
201
A
210
211
212
220
202
203
220
B
C
204
205
206
207
D
E
F
G
230
230
230
230
160

Fig. 2

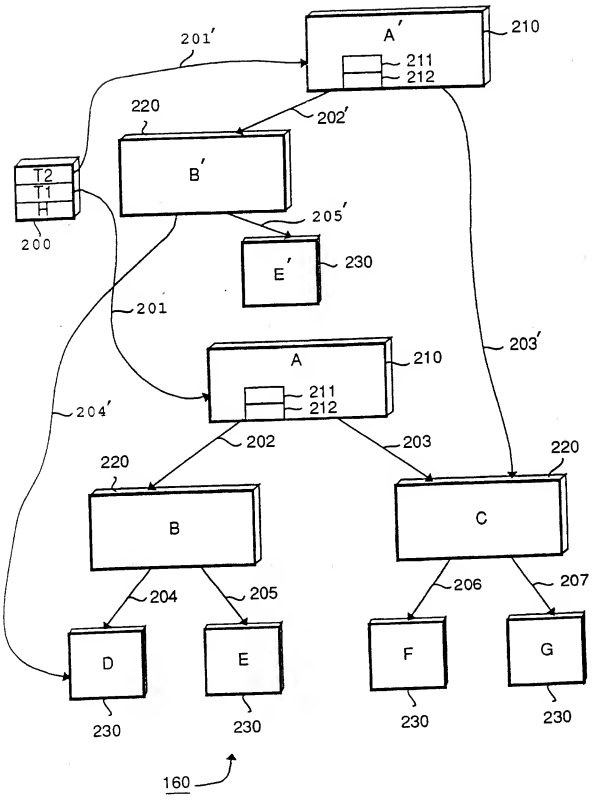
A'
210
211
212
201'
202'
220
B'
T2
T1
H
200
205'
E'
230
201
203'
A
210
211
212
204'
202
203
220
220
B
C
204
205
206
207
D
E
F
G
230
230
230
230
160

Fig. 3

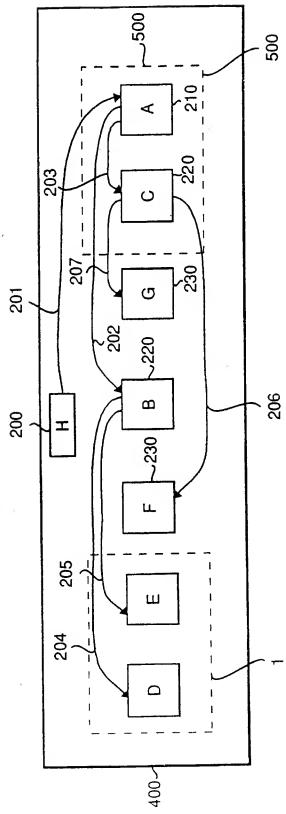
500
500
A
210
203
C
220
207
G
230
201
202
B
220
200
H
206
230
F
205
E
204
D
1
400

Fig. 4

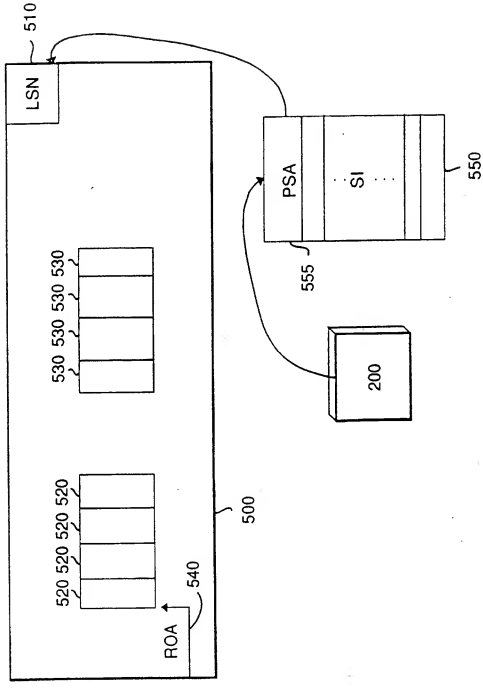
510
LSN
530 530 530 530
520 520 520 520
540
ROA
500
PSA
SI
550
555
200

Fig. 5

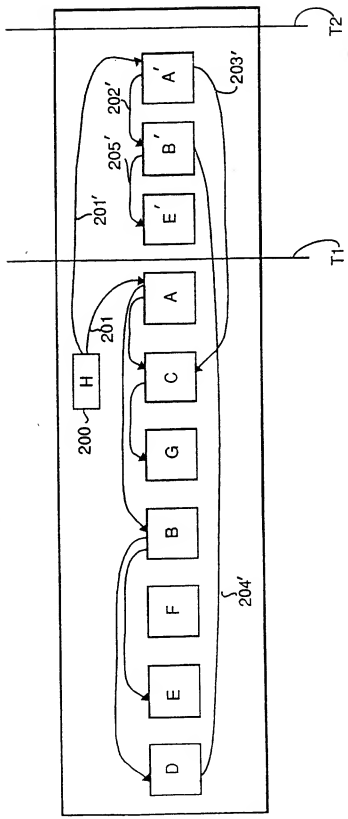

Fig. 6

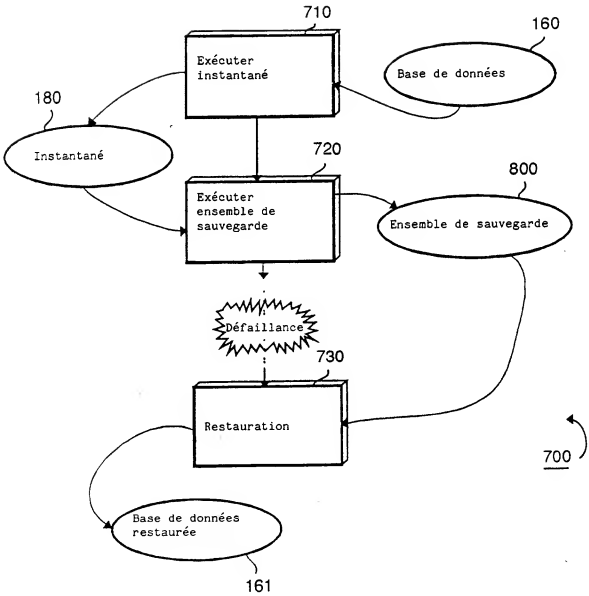
710
Exécuter instantané
160
Base de données
180
Instantané
720
Exécuter ensemble de sauvegarde
800
Ensemble de sauvegarde
Défaillance
730
Restauration
700
Base de données restaurée
161

Fig. 7

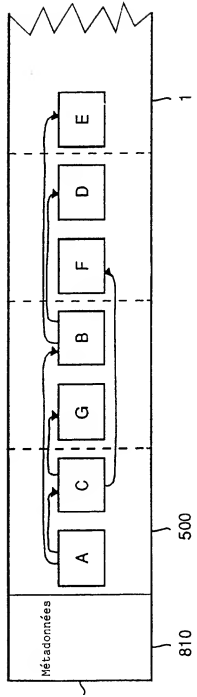

Fig. 8

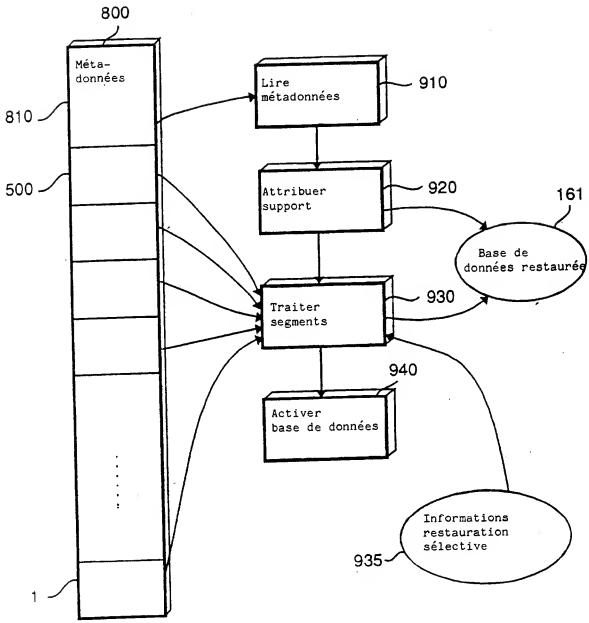
800
Méta-
données
810
500
1
Lire
métadonnées
910
Attribuer
support
920
161
Base de
données restaurée
Traiter
segments
930
940
Activer
base de données
Informations
restauration
sélective
935

Fig. 9

REPUBLIQUE FRANÇAISE

INSTITUT NATIONAL
de la
PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE PRELIMINAIRE

établi sur la base des dernières revendications déposées avant le commencement de la recherche

2746526

N° d'enregistrement national

FA 525843
FR 9603690

DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendications concernées de la demande examinée
A	US-A-5 119 490 (KUROSE HIROYASU) 2 Juin 1992 * abrégé * -----	1

DOMAINES TECHNIQUES RECHERCHES (Int.Cl.6)

G06F

1

Date d'achèvement de la recherche	Examinateur
13 Décembre 1996	Katerbau, R

CATEGORIE DES DOCUMENTS CITES

X : particulièrement pertinent à lui seul
Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie
A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général
O : divulgation non-écrite
P : document intercalaire

T : théorie ou principe à la base de l'invention
E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.
D : cité dans la demande
L : cité pour d'autres raisons

& : membre de la même famille, document correspondant

FPO FORM 1503 03.82 (P04C13)